DOCUMENT RESUME

ED 235 222 TM 830 643

AUTHOR Bart, William M.; Palvia, Rajkumari
TITLE Relationships among Test Factor Structure, Test Hierarchical Structure, and Test Inter-Item Dependency Structure.
SPONS AGENCY Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
PUB DATE Apr 83
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Factor Structure; *Interaction; *Latent Trait Theory; Mathematical Models; Multivariate Analysis; Statistical Analysis; *Test Items
IDENTIFIERS *Hierarchical Analysis; Log Linear Models

ABSTRACT

In previous research, no relationship was found between test factor structure and test hierarchical structure. This study found some correspondence between test factor structure and test inter-item dependency structure, as measured by a log-linear model. There was an inconsistency, however, which warrants further study: more significant two-item interactions were found for the set of items which had been found to be most heterogeneous by latent trait techniques, rather than for the more homogeneous set of items. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235222

Relationships among Test Factor Structure, Test Hierarchical Structure, and Test Inter-Item Dependency Structure

by

William M. Bart

and

Rajkumari Palvia

University of Minnesota

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. M. Bart

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 653

Paper presented at the American Educational Research Association meeting, Montreal, 1983.

Abstract

In previous research, no relationship was found between test factor structure and test hierarchical structure. This study found some correspondence between test factor structure and test inter-item dependency structure, as measured by a log-linear model. There was an inconsistency, however, which warrants further study: more significant two-item interactions were found for the set of items which had been found to be most heterogeneous by latent trait techniques, rather than for the more homogeneous set of items.

Relationships among Test Factor Structure, Test Hierarchical Structure, and Test Inter-Item Dependency Structure

The analysis of structure among dichotomously scored items in a test continues to be a central concern in psychometrics. The structure in a test has been conceived in two standard ways: (a) test factor structure, which is the complex of continuous latent traits (factors or dimensions) which the items assess and (b) test hierarchical structure, which is the system of prerequisite relations among the discrete items. Latent trait test theory has been used to investigate test factor structure (e.g., Lord & Novick, 1968; Bock & Lieberman, 1970) and ordering analysis has been used to investigate test hierarchical structure (Airasian & Bart, 1973; Bart & Airasian, 1974; Bart & Krus, 1973).

Bart (1978) examined the relationship between those two types of test structure. He started with data for two sets of items such that one set of items were established to have a unifactor structure and the other set of items a two-factor structure. These sets of item data were then analyzed with a refined ordering-analytic method to determine item hierarchies. No hierarchy structure was found for either set of items. Thus, Bart suggested that test factor structure provides little, if any, information about test hierarchical structure.

A third type of structure is the test inter-item dependency structure, which is the array of statistical interdependencies among the items. Discrete multivariate analysis with its log-linear modeling has been developed to investigate statistical interdependencies among discrete variables and can be used to investigate test inter-item dependency structure (Bishop,

Fienberg, & Holland, 1975). This study is intended to investigate the inter-item dependency structures of two tests whose test factor structures and test hierarchical structures have already been determined.

Method

Basic to the log-linear model is the cross-product ratio, α. Please consider the case of two dichotomously-scored items, A_1 and A_2, for which a 2 x 2 table can be constructed depicting four cell probabilities, p_{00}, p_{01}, p_{10}, and p_{11}, such that p_{ij} (i=0,1; j=0,1) is the probability of receiving a score of i for item A_1 and a score of j for item A_2. A score of 1 indicates success on the item and a score of 0 indicates failure with the item.

[Insert Table 1 about here]

Using the 2 x 2 table cited in Table 1, one can define the cross-product ratio as

$$\alpha = \frac{p_{00} \cdot p_{11}}{p_{10} \cdot p_{01}} . \qquad (1)$$

The logarithm of the cross-product ratio, α, is a linear contrast of the log-probabilities of the four cells, namely

$$\log \alpha = \log p_{00} - \log p_{10} - \log p_{10} + \log p_{11} . \qquad (2)$$

When $\log \alpha = 0$, items A_1 and A_2 are said to be statistically independent. If $\log \alpha > 0$, items A_1 and A_2 are said to be statistically dependent.

To articulate certain elemental features of the log-linear model, we

will closely follow the exposition provided by Bishop, Fienberg, and Holland (1975). Proceeding in a manner analogous to analysis of variance (ANOVA) models, a linear model can be constructed with the use of the natural logarithms of the cell probabilities. One could then write

$$\log p_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad i=0,1;\ j=0,1 \qquad (3)$$

where u is the grand mean of the natural logarithms of the probabilities:

$$u = 1/4\ (\log p_{00} + \log p_{01} + \log p_{10} + \log p_{11}), \qquad (4)$$

$u + u_{1(i)}$ is then expressed as the mean of the natural logarithms of the probabilities at level i of the first item,

$$u + u_{1(i)} = 1/2\ (\log p_{i0} + \log p_{i1}), \qquad i = 0,1, \qquad (5)$$

and $u + u_{2(j)}$ is expressed as the mean of the logarithms of the probabilities at level j of the second item,

$$u + u_{2(j)} = 1/2\ (\log p_{0j} + \log p_{1j}) \qquad j = 0,1. \qquad (6)$$

Both $u_{1(i)}$ and $u_{2(j)}$ represent deviations from the grand mean u,

$$u_{1(0)} + u_{1(1)} = u_{2(0)} + u_{2(1)} = 0, \qquad (7)$$

and $u_{12}(ij)$ represents a deviation from $u + u_{1(i)} + u_{2(j)}$, so that

$$u_{12(11)} = -u_{12(01)} = -u_{12(10)} = u_{12(11)}. \qquad (8)$$

If one defines $\ell_{ij} = \log p_{ij}$, then one can then write the grand mean as

$$u = \frac{\ell_{++}}{4} = \frac{\sum_{ij} \ell_{ij}}{4} = \frac{\ell_{00} + \ell_{01} + \ell_{10} + \ell_{11}}{4} \qquad (9)$$

with the main effects being

$$u_{1(i)} = \frac{\ell_{i+}}{2} - \frac{\ell_{++}}{4} \qquad (10)$$

and

$$u_{2(j)} = \frac{\lambda_{+j}}{2} - \frac{\lambda_{++}}{4} \qquad (11)$$

and the interaction term being

$$u_{12(ij)} = \lambda_{ij} - \frac{\lambda_{i+}}{2} - \frac{\lambda_{+j}}{2} + \frac{\lambda_{++}}{4} . \qquad (12)$$

Bishop, Fienberg, & Holland (ibid.) close this introductory section by pointing out that "the main effects are functions of the marginal sums of the logarithms but do not correspond to the marginal sums p_{i+} and p_{+j} in the original scale." Thus item probability logarithms and their marginal sums become very important in this analysis.

The log-linear model for two items is analogous to a two-way ANOVA. Log-linear models for n items than are analogous to n-way ANOVA's. This study is an investigation of the interdependencies among two sets of five items and thus employs a log-linear model comparable to a five-way ANOVA.

Either the items in each test data set are completely independent of each other or any two or more items are interdependent. The data for this study consists of items 11 through 15 of Section 6 (Figure Classification) and Section 7 (Debate) of the Law School Admission Test and was the same data analyzed by Bock & Lieberman (1970), Christoffersson (1975), and Bart (1978). To represent the two data sets, the responses of 1000 subjects drawn from a larger sample of students applying for admission to Law School at various universities in the U.S.A. are recorded as binary responses with "1" indicating a correct item response and "0" indicating an incorrect item response. Thus there are $2^5 = 32$ possible item response patterns for each

section of the test. The frequencies of the various item response patterns are recorded in Table 2.

[Insert Table 2 about here]

The data for both sets of items can be called as LSAT6 and LSAT7 respectively and can be interpreted as constituting five-dimension contingency tables with each item representing a discrete variable and each item score as a variable category. Since the responses are binary for each item, there are only two categories for each item variable and we have a 2 x 2 x 2 x 2 x 2 contingency table.

Let $\log p_{ijklm}$ denote the logarithm of the probability of the (i,j,k,l,m) cell in a five-way contingency cell. Then a "saturated" log-linear model for the data can be written as

$$\begin{aligned}\log p_{ijklm} = {} & u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)} \\ & + u_{12(ij)} + u_{13(ik)} + \cdots + u_{35(km)} + u_{45(lm)} \\ & + u_{123(ijk)} + \cdots + u_{345(klm)} \\ & + u_{1234(ijkl)} + \cdots + u_{2345(jklm)} \\ & + u_{12345(ijklm)} \\ & \text{with } i=0,1;\ j=0,1;\ k=0,1;\ l=0,1;\ m=0,1. \qquad (13)\end{aligned}$$

The u-terms $u_1, u_2, u_3, \ldots, u_{12345}$ are parameters of the model and have the usual meaning as in a general log-linear model. (For more details, one can consult Bishop, Fienberg, & Holland (1975)). The number of independent parameters in the saturated model being used is 32 which is the number of cells in the five-way contingency table representing each data set.

Results

Models derived from the five variable saturated model were used in this analysis. To be specific, only hierarchical models were fitted to the data which means that, if any of the u-terms is set to zero, all of its higher-order relatives must also be zero. For example, if $u_{12} = 0$, then u_{123}, u_{124}, u_{125}, u_{1234}, u_{1235}, u_{1245}, and u_{12345} must also be zero. (Bishop, Fienberg, & Holland, 1975).

The first model to be examined is the model of independence which can be represented as

$$\log p_{ijklm} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)}. \quad (14)$$

This model was fitted to the two sets of data and chi-square values for both sets were quite large. The chi-square value for the LSAT6 (Figure Classification) data was 95.06 with a p-value $<$ .005 and the chi-square value for the LSAT7 (Debate) data was 356.6 with a p-value $<$ .005. Hence, the hypothesis of independence among the test items was rejected for both sets of items, suggesting that one or more interaction terms are present within the LSAT6 as well as LSAT7 item sets.

The next step in the analysis was to find the interaction(s) present in each set of data (i.e., the interactions that were significantly different from zero). The model with all two factor interactions present, but all higher-order interaction terms being zero fitted very well each of the two data sets. This result is quite important in that it established that there were no complex three or four or five factor interactions present among the items of both sets of data and the presence of significant second-order interactions can be compared directly with the inter-item correlations presented by Bock & Lieberman (1970). This comparison is discussed at

greater length after a few more results are provided.

The next step was to see which of the two-factor interaction terms can be further eliminated from the model of all two-factor interactions (with no higher-order interactions being present) without significantly lowering the p-value.

[Insert Tables 3 and 4 about here]

The ten models with one two-factor interaction term being removed for each model were fitted to each set of test item data. Tables 3 and 4 depict the most prominent results of these analyses for LSAT6 and LSAT7 data sets respectively. For LSAT6 data, the $u_{15} = 0$ model fitted the best of all ten models without even affecting the p-value. Having u_{15} eliminated, there were only nine two factor interaction terms in the model. As was done previously, nine models were fitted with one interaction term (besides u_{15}) being removed at a time. The term u_{14} was next to be eliminated from the original model. The procedure was repeated until no more interaction terms could be removed. This same procedure was repeated for LSAT7 data.

The selection of the best model depends on the p-value set by the experimenter or by the purpose of the study. If, through the elimination of an interaction term, there is a sudden drop in p-value, then one may want to maintain that interaction term in the model. Using that principal in the fitting of the various interaction models, the best model for the LSAT6 data was found to be

$$\begin{aligned}\log p_{ijk\,m} = {} & u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)} \\ & + u_{13(ik)} + u_{23(jk)} + u_{25(jm)} \\ & + u_{34(kl)} + u_{45(lm)} \end{aligned} \qquad (15)$$

and this model has only 11 independent parameters as compared to the original model. The best model describing the LSAT7 data was found to be

$$\log p_{ijk\,m} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)} + u_{12(ij)} + u_{13(ik)} + u_{14(il)} + u_{15(im)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)} + u_{35(km)} \qquad (16)$$

which has 14 independent parameters.

For the LSAT6 data, the only interaction term present in the $u_{12} = u_{14} = u_{15} = 0$ model relating to item 1 was u_{13}. This implies that item 1 was related to only item 3 and is conditionally independent of all of the other items. Since u_{13}, u_{23}, and u_{34} were present in the model but $u_{35} = 0$, item 3 was found to be related to all of the items except item 5.

For the LSAT7 data, all interaction terms containing item 1 and 3 such as u_{12}, u_{13}, u_{14}, u_{15}, u_{23}, u_{34}, and u_{35} were significant and thus present. This condition implied that item 1 was related to all of the other items and the same was the case with item 3. Since $u_{25} = u_{45} = 0$, item 2 was related to all of the other items and the same was the case with item 3. Since $u_{25} = u_{45} = 0$, item 2 was related to all of the other items except item 5. Item 4 was also related to all of the other items except item 5.

The results of this study seem to be in accordance with the results regarding the tetrachoric correlation coefficients reported by Bock & Lieberman (1970). For the LSAT6 data, the interaction terms of u_{14}, u_{15}, u_{24}, and u_{35} were the first four interaction terms to be eliminated from the model of all two-factor interactions; this situation was reflected in the fact that the correlation coefficients for item pairs (1,4), (1,5), (2,4), and (3,5) were the smallest inter-item correlations cited for the

LSAT7 data by Bock & Lieberman (ibid.).

One result emanating from this study which conflicted with results emanating from the studies by Bock & Lieberman (1970) and Christoffersson (1975) regarded the heterogeneity of the items. They contended that the LSAT7 items were more heterogeneous than the LSAT6 items, but there were larger inter-item tetrachoric correlation coefficients among the LSAT7 items and there were more significant interaction terms present in the best-fitting log-linear model for the LSAT6 data. Despite the results of the latent trait test theoretic analyses performed on the data in those two studies, the LSAT7 items may have been much less heterogeneous than reported by those studies. A closer examination of the similarities and differences among the items is needed to identify item features which may contribute to the homogeneity among items.

Discussion

Bart (1978) established that both sets of test items had (a) no hierarchical structures, (b) prominent first linear order but no clear uniorder or two-order structures, and (c) internal inconsistency. He determined empirically that factor structures and hierarchical structures cannot be simply mapped into each other. In this study, an array of significant two-item interactions were located for both test item sets with more significant two-item interactions being located for the set of items found to be most heterogeneous by latent trait techniques. This inconsistency warrants further examination before an adequate explanation can be generated. One would assume that there would be substantial correspondence between the test factor structure and the test inter-item dependency structure, but the

results of this study indicated that the correspondence is not a strong one. It is suggested that all three types of test structure be investigated for any test item data set and that the instructional consequences of each of the three structures be investigated.

References

Airisian, P. & Bart, W. Ordering theory: a new and useful measurement model. Educational Technology, 1973, 18, 36-60.

Bart, W. An empirical inquiry into the relationship between test factor structure and test hierarchical structure. Applied Psychological Measurement, 1978, 2, 333-337.

Bart, W. & Airasian, P. The determination of the ordering among seven Piagetian tasks by an ordering-theoretic method. Journal of Educational Psychology, 1974, 9, 251-259.

Bart, W. & Krus, D. An ordering-theoretic method to determine hierarchies among items. Educational and Psychological Measurement, 1973, 33, 291-300.

Bishop, Y., Fienberg, S., and Holland, P. Discrete multivariance analysis: theory and practice. Cambridge, Massachusetts: MIT Press, 1975.

Bock, R. & Lieberman, M. Fitting a response model for a dichotomously scored item. Psychometrika, 1970, 35, 179-197.

Christoffersson, A. Factor analysis of dichotomized variables. Psychometrika, 1975, 40, 5-32.

Lord, F. & Novick, M. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

Footnote

The research reported herein and the preparation of this paper was supported in part by grants to the author from the United States Army Research Institute for the Behavioral and Social Sciences. Preparation of this paper was also supported in part by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB-35703X), the National Institute of Child Health and Human Development (HD-00136), and the Graduate School of the University of Minnesota.

Requests for additional copies should be sent to William M. Bart, 330 Burton Hall, University of Minnesota, 178 Pillsbury Drive S.E., Minneapolis, Minnesota 55455.

Table 1

An Inter-item Contingency Table

		A_2 = 0	A_2 = 1
A_1	0	P_{00}	P_{01}
	1	P_{10}	P_{11}

Table 2

Frequencies of Item Response Patterns

For Two Sets of Law School Admissions Test Items

Index	Response Pattern of Items					LSAT6 Figure Classification	LSAT7 Debate
	1	2	3	4	5	Frequency	Frequency
1	0	0	0	0	0	3	12
2	1	0	0	0	0	10	7
3	0	1	0	0	0	1	10
4	1	1	0	0	0	16	6
5	0	0	1	0	0	1	3
6	1	0	1	0	0	3	14
7	0	1	1	0	0	0	7
8	1	1	1	0	0	11	18
9	0	0	0	1	0	2	1
10	1	0	0	1	0	14	11
11	0	1	0	1	0	0	3
12	1	1	0	1	0	21	7
13	0	0	1	1	0	3	3
14	1	0	1	1	0	15	15
15	0	1	1	1	0	2	8
16	1	1	1	1	0	28	32
17	0	0	0	0	1	6	19
18	1	0	0	0	1	29	39
19	0	1	0	0	1	8	5
20	1	1	0	0	1	56	25
21	0	0	1	0	1	1	19
22	1	0	1	0	1	28	51
23	0	1	1	0	1	3	23
24	1	1	1	0	1	61	136
25	0	0	0	1	1	11	7
26	1	0	0	1	1	81	34
27	0	1	0	1	1	16	7
28	1	1	0	1	1	173	35
29	0	0	1	1	1	4	17
30	1	0	1	1	1	80	90
31	0	1	1	1	1	15	28
32	1	1	1	1	1	298	308

Table 3

The Fit of Each of Several Log-Linear Models

for a Set of Five Law School Admissions Test (LSAT6) Items

Assessing Figure Classification Ability

Model		G^2	X^2	df.	p-value
(1)	Independence with all interactions absent.	74.80	95.06	26	$<.005$
(2)	All three factor and higher-order interactions absent.	16.48	14.65	16	$>.25$
(3)	$u_{15} = 0$	16.59	15.34	17	$>.25$
(4)	$u_{15} = u_{14} = 0$	17.45	16.46	18	$>.25$
(5)	$u_{15} = u_{14} = u_{35} = 0$	18.62	17.11	19	$>.25$
(6)	$u_{15} = u_{14} = u_{35} = u_{24} = 0$	20.36	19.38	20	$>.25$
(7)	$u_{15} = u_{14} = u_{35} = u_{24} = u_{12} = 0$	24.06	24.95	21	$\sim.25$
(8)	$u_{12} = u_{14} = u_{15} = u_{24} = u_{25} = u_{35} = 0$	30.93	34.96	22	$=.1$
(9)	$u_{12} = u_{14} = u_{15} = u_{24} = u_{25} = u_{34} = u_{35} = 0$	42.78	47.74	23	$<.001$
(10)	$u_{12} = u_{14} = u_{15} = u_{23} = u_{24} = u_{25} = u_{34} = u_{35} = 0$	236.1	229.0	24	$<.005$

Table 4

The Fit of Each of Several Log-linear Models

for a Set of Five Law School Admissions Test (LSAT7) Items

Assessing Debate Ability

Model	G^2	X^2	df.	p-value
(1) Independence with all interactions absent.	200.9	356.6	26	$\ll .005$
(2) All three factor and higher order interactions absent.	20.38	20.14	16	$.2 < p < .25$
(3) $u_{25} = 0$	20.86	20.85	17	$.2 < p < .25$
(4) $u_{45} = u_{25} = 0$	23.02	23.50	18	$.1 < p < .2$
(5) $u_{12} = u_{45} = u_{25} = 0$	28.40	28.54	19	$.05 < p < .1$
(6) $u_{24} = u_{45} = u_{25} = 0$	28.82	29.78	19	$.05 < p < .1$
(7) $u_{24} = u_{12} = u_{45} = u_{35} = o$	36.05	36.61	20	$.01 < p < .05$